Combinatorics, Probability and Computing (1998) 7, 221–238. Printed in the United Kingdom

Ramsey Theory and Sequences of Random Variables

WILLIAM T. TROTTER[1]† and PETER WINKLER[2]†

[1] Department of Mathematics, Arizona State University,
Tempe, AZ 85287, USA
(e-mail: trotter@ASU.edu)

[2] Bell Laboratories 2C-379, Murray Hill, NJ 07974, USA
(e-mail: pw@bell-labs.com)

Received 19 September 1996; revised 24 March 1997

We consider probability spaces which contain a family $\{E_A : A \subseteq \{1, 2, \ldots, n\}, |A| = k\}$ of events indexed by the k-element subsets of $\{1, 2, \ldots, n\}$. A pair (A, B) of k-element subsets of $\{1, 2, \ldots, n\}$ is called a *shift pair* if the largest $k - 1$ elements of A coincide with the smallest $k - 1$ elements of B. For a shift pair (A, B), $\Pr[A\overline{B}]$ is the probability that event E_A is true and E_B is false. We investigate how large the minimum value of $\Pr[A\overline{B}]$, taken over all shift pairs, can be. As $n \to \infty$, this value converges to a number λ_k, with $\frac{1}{2} - \frac{1}{2k+2} \leqslant \lambda_k \leqslant \frac{1}{2} - \frac{1}{4k+2}$. We show that λ_k is a strictly increasing function of k, with $\lambda_1 = \frac{1}{4}$ and $\lambda_2 = \frac{1}{3}$.

For $k = 1$, our results have the following natural interpretation. If a fair coin is tossed repeatedly, and event E_i is true when the ith toss is heads, then for all i and j with $i < j$, $\Pr[E_i\overline{E_j}] = \frac{1}{4}$. Furthermore, as we show in this paper, for any $\varepsilon > 0$, there is an n such that for any sequence $E_1, E_2, \ldots, E_n$ of events in an arbitrary probability space, there are indices $i < j$ with $\Pr[E_i\overline{E_j}] < \frac{1}{4} + \varepsilon$. The results and techniques we develop in this research, together with further applications of Ramsey theory, are then used to show that the supremum of fractional dimensions of interval orders is exactly 4, answering a question of Brightwell and Scheinerman.

Generalizing the $\frac{1}{4} + \varepsilon$ result to random variables $X_1, X_2, \ldots, X_n$ with values in an m-element set, we obtain a finite version of de Finetti's theorem without the exchangeability hypothesis: for any fixed m, k and ε, every sufficiently long sequence of such random variables has a length-k subsequence at variation distance less than ε from an i.i.d. mix.

† Research supported in part by the Office of Naval Research.

1. Introduction

The use of Ramsey theory in the study of sequences (especially infinite sequences) of random variables is not new; see, for instance, [1], [3], [4], [7] and [13] for applications related to the 'subsequence principle' of probability theory. Our application will be slightly different in flavour and limited (on account of motivation) to finite sequences; however, our results extend to infinite sequences in a straightforward manner.

We begin with an elementary example and a follow-up question which serve to motivate much of the material to follow. First, the example. Suppose a fair coin is tossed n times. Define event E_i to be true when the ith toss is 'heads'. Then, for all i, j with $1 \leqslant i < j \leqslant n$, $\Pr[E_i\overline{E_j}] = \frac{1}{4}$. If we condition on exactly $n/2$ heads (n even), we can increase $\Pr[E_i\overline{E_j}]$ slightly to $\frac{1}{4} \cdot \frac{n}{n-1}$. Now the question. Can we do *asymptotically* better? We can state this more formally as follows.

Question 1.1. *Does there exist a number $\lambda > 1/4$, so that for every $n \geqslant 2$ there exists a probability space with events $E_1, E_2, \ldots, E_n$ such that $\Pr[E_i\overline{E_j}] \geqslant \lambda$, for all $i < j$?* □

We will show that the answer to Question 1.1 is 'no': any sufficiently long sequence of events in any probability space contains a subsequence which does little better than flipping a fair coin when it comes to keeping $\Pr[E_i\overline{E_j}]$ large. The Ramsey-theoretic flavour of this statement, together with its quantitative hedging, is characteristic of the theorems to follow.

In order to discuss generalizations of this elementary example, we need some additional notation and terminology. First, we must extend the notion of sequences of events to families of events indexed by subsets of a finite set.

For a positive integer n, we let $[n]$ denote the n-element set $\{1, 2, \ldots, n\}$. When S is a finite set and $0 \leqslant k \leqslant |S|$, we let $\binom{S}{k}$ denote the family of all k-element subsets of S. Given a probability space Ω, a finite set S of positive integers, and an integer k with $0 \leqslant k \leqslant |S|$, a (k, S)-*scheme* in Ω is just a family $\mathscr{S} = \{E_A : A \in \binom{S}{k}\}$ of events from Ω indexed by the k-element subsets of S. When $\mathscr{S}$ is a (k, S) scheme and $T \subseteq S$, the family $\mathscr{T} = \{E_A : A \in \binom{T}{k}\}$ is called a *subscheme* of $\mathscr{S}$. We also say $\mathscr{T}$ is the subscheme *determined* by T.

When $\mathscr{S} = \{E_A : A \in \binom{S}{k}\}$ is a (k, S)-scheme, we abuse notation slightly and blur the distinction between the subset A and the event E_A. So, we refer to a set $A \in \binom{S}{k}$ as an event in Ω and write $\Pr[A]$ rather than $\Pr[E_A]$. Also, we write $\Pr[AB]$ rather than $\Pr[E_AE_B]$ and $\Pr[A\overline{B}]$ rather than $\Pr[E_A\overline{E_B}]$. When $k = 1$, we write $\Pr[i]$ rather than $\Pr[\{i\}]$, *etc.*

When listing the elements of a finite set of integers, we will *always* list them in increasing order; for instance, the statement $\{i, j, k\} \in \binom{[n]}{3}$ also implies that $i < j < k$. When S is a finite set and $A, B \in \binom{S}{k}$, the ordered pair (A, B) is called a (k, S)-*shift pair* when there is a subset $\{i_1, i_2, \ldots, i_{k+1}\} \subseteq S$ so that $A = \{i_1, i_2, \ldots, i_k\}$ and $B = \{i_2, i_3, \ldots, i_{k+1}\}$. For emphasis, we point out that our notational conventions imply that $i_1 < i_2 < \cdots < i_{k+1}$ in this definition.

Now fix an integer $k \geqslant 1$ and let S be a set with $|S| > k$. When $\mathscr{S}$ is a (k, S)-scheme, we let $\lambda(\mathscr{S}) = \min\{\Pr[A\overline{B}] : (A, B)$ is a (k, S)-shift pair$\}$. In turn, we set $\lambda(k, n)$ to be the maximum value of $\lambda(\mathscr{S})$, taken over all probability spaces and all (k, S)-schemes with $|S| = n$. From its definition, $\lambda(k, n)$ is a decreasing function of n, so we may define $\lambda_k = \lim_{n\to\infty} \lambda(k, n)$. In particular, the negative answer to Question 1.1 will follow from the assertion of Theorem 3.1 that $\lambda_1 = 1/4$.

We will show that λ_k is a strictly increasing function of k and satisfies the following bounds:

$$\frac{1}{2} - \frac{1}{2k+2} \leqslant \lambda_k \leqslant \frac{1}{2} - \frac{1}{4k+2}.$$

Originally, we guessed that the lower bound in this inequality was tight. If this conjecture were true, it would imply that $\lambda_5 = 5/12$, but we have been able to prove that $\lambda_5 \geqslant 27/64 > 5/12$. So we are now hesitant to hazard a guess for the form of $f(k)$ as a function of k. On the other hand, we will prove that $\lambda_2 = 1/3$, and we believe that $\lambda_3 = 3/8$ and $\lambda_4 = 2/5$.

The remainder of the paper is organized as follows. In Section 2, we use Ramsey theory to develop concepts of regularity and uniformity for schemes. These concepts are then used in Section 3 to show that $\lambda_1 = \frac{1}{4}$ and $\lambda_2 = \frac{1}{3}$. In Sections 4 and 5, we provide bounds for λ_k. In Sections 6 and 7, we discuss the chromatic number of shift graphs and the motivating problem from fractional dimension theory for partially ordered sets. Using the techniques developed in this paper, we then solve this problem.

Finally, in Section 8, we return to Question 1.1 and generalize to random variables, obtaining a new finite form of de Finetti's theorem.

2. Regularity and uniformity

Let $\mathscr{S}$ be a (k, S)-scheme in a probability space Ω. There is no reason why any two events in $\mathscr{S}$ should have exactly the same probability. However, if $|S|$ is sufficiently large in comparison to k, it seems reasonable that there should be a large subset $U \subseteq S$ so that any two events in the subscheme determined by U have approximately the same probability. To formalize this notion, let $\varepsilon > 0$. Given an event E in Ω, there is a unique integer i so that

$$i\varepsilon \leqslant \Pr[E] < (i+1)\varepsilon.$$

We may then use the value $i\varepsilon$ as an approximation for the probability of E. Note that the number of distinct values used in approximating the probabilities of events in Ω is $1 + \lfloor \frac{1}{\varepsilon} \rfloor$, which depends only on ε. The following result is then an immediate consequence of Ramsey's theorem.

Proposition 2.1. *For positive integers k and n, with $k < n$, and a real number $\varepsilon > 0$, there is an integer n_0 so that if $\mathscr{S}$ is a (k, S)-scheme in a probability space Ω with $|S| \geqslant n_0$, then there exists a subset $U \subseteq S$ with $|U| = n$ so that $|\Pr[A] - \Pr[A']| < \varepsilon$, for every $A, A' \in \binom{U}{k}$.* □

However, we will find it useful to work with much stronger notions of regularity. Let k and n be positive integers, with $k < n$, and let $\mathscr{S}$ be a (k, S)-scheme with $|S| = n$.

Now let s be a positive integer with $k \leqslant s \leqslant n$. To a subset $R = \{i_1, \ldots, i_s\} \in \binom{S}{s}$, we can associate a random variable $\mathbf{X}_R$ whose values are subsets of $\binom{[s]}{k}$; namely, $\mathbf{X}_R = \{\{j(1), \ldots, j(k)\}$: the event $\{i_{j(1)}, \ldots, i_{j(k)}\}$ is true$\}$. Any particular subset P of $\binom{[s]}{k}$ will be called a *(k,s)-pattern*. Given $\varepsilon \geqslant 0$, we say that $\mathscr{S}$ is *ε-regular* when, for every s with $k \leqslant s \leqslant n$ and every pair $R, R' \in \binom{S}{s}$,

$$|\Pr[\mathbf{X}_R = P] - \Pr[\mathbf{X}_{R'} = P]| < \varepsilon,$$

for every (k,s)-pattern P. The following theorem is an immediate consequence of Ramsey's theorem.

Proposition 2.2. *For positive integers k and n, with $k < n$, and a real number $\varepsilon > 0$, there is an integer n_0 such that if $\mathscr{S}$ is any (k,S)-scheme with $|S| \geqslant n_0$, then there is a subset $U \subseteq S$ with $|U| = n$ which determines a (k,U)-subscheme $\mathscr{U}$ which is ε-regular.*

A 0-regular scheme is said to be *regular*. Using Ramsey theory alone, it does not appear possible to deduce the existence of regular subschemes from parent schemes which are merely ε-regular for some $\varepsilon > 0$. However, with a little more analysis, we find that we can construct regular schemes by passing to the limit.

Theorem 2.3. *For positive integers k and n, with $k < n$, there exists a probability space Ω and a $(k,[n])$-scheme $\mathscr{S}$ in Ω so that*

(1) *$\mathscr{S}$ is regular, and*
(2) *$\lambda(\mathscr{S}) \geqslant \lambda_k$.*

Proof. Let k and n be positive integers with $k < n$. For each positive integer i, set $\varepsilon_i = \frac{1}{i}$. Apply Proposition 2.2 and the definition of λ_k to select a sequence of probability spaces $\{\Omega_i : i \geqslant 1\}$ so that Ω_i contains a $(k,[n])$-scheme $\mathscr{S}_i$ which is ε_i-regular and satisfies $\lambda(\mathscr{S}_i) \geqslant \lambda_k$.

For each $i \geqslant 1$, each s with $k \leqslant s \leqslant n$, and each (k,s)-pattern P, let $p(i,P) = \Pr[\mathbf{X}_{[s]} = P]$ in Ω_i. For each $i \geqslant 1$, the number of patterns is $\sum_{s=k}^{n} 2^{\binom{s}{k}}$, which is of course bounded as a function of k and n. It follows that we may choose a subsequence $\{\Omega_{i_j} : j \geqslant 1\}$ for which each of the sequences $\{p(i_j,P) : j \geqslant 1\}$ converges, say to a value $p(P)$.

Finally, we define the probability space Ω in the obvious way. The elementary events in Ω correspond to the $(k,[n])$ patterns of $[n]$, that is, the elementary events are just the subsets of $\binom{[n]}{k}$. In Ω, we take the probability of the pattern P to be $p(P)$. This definition determines a regular $(k,[n])$-scheme $\mathscr{S}$ with $\lambda(\mathscr{S}) \geqslant \lambda_k$. □

The preceding result allows us to make several additional assumptions about the probability spaces we consider in determining λ_k. For starters, note that, if $\mathscr{S}$ is a regular (k,S)-scheme, and (A,B) is a (k,S)-shift pair, then $\Pr[A\overline{B}] = \lambda(\mathscr{S})$. Moreover, since $\Pr[A] = \Pr[B]$, $\Pr[A] = \Pr[AB] + \Pr[A\overline{B}]$, and $\Pr[B] = \Pr[AB] + \Pr[\overline{A}B]$, it follows that $\Pr[A\overline{B}] = \Pr[\overline{A}B]$. As a consequence, the meanings of 'true' and 'false' may be reversed. To see this, note that we define a new probability space Ω' containing a regular (k,S)-scheme $\mathscr{S}'$ as follows. Flip a fair coin. If the toss is heads, set A to be true in Ω' if and only if A

is true in Ω. If the toss is tails, set A to be true in Ω' if and only if A is false in Ω. In the new space Ω', $\Pr[A] = \frac{1}{2}$, for all $A \in \binom{S}{k}$. Furthermore, $\lambda(\mathscr{S}) = \lambda(\mathscr{S}')$.

We can extend these concepts to larger sets as follows. Denote by P^* the *reverse* of the pattern P, so that $\{j(1), j(2), \ldots, j(k)\} \in P^*$ just when $\{s{+}1{-}j(k), s{+}1{-}j(k{-}1), \ldots, s{+}1{-}j(1)\} \in P$. Also, let $\overline{P} = \binom{[s]}{k} \setminus P$ denote the *complement* of P. We say that a (k, S)-scheme $\mathscr{S}$ is *uniform* when for every s with $k \leqslant s \leqslant |S|$ every pair $R, R' \in \binom{S}{s}$, and every (k, s)-pattern P,

$$\begin{aligned}\Pr[\mathbf{X}_R = P] &= \Pr[\mathbf{X}_{R'} = P],\\ \Pr[\mathbf{X}_R = P] &= \Pr[\mathbf{X}_{R'} = P^*], \quad \text{and}\\ \Pr[\mathbf{X}_R = P] &= \Pr[\mathbf{X}_{R'} = \overline{P}].\end{aligned}$$

Theorem 2.4. *For positive integers k and n, with $k < n$, there exists a probability space Ω and a $(k, [n])$-scheme $\mathscr{S}$ in Ω so that*

(1) *$\mathscr{S}$ is uniform, and*
(2) *$\lambda(\mathscr{S}) \geqslant \lambda_k$.*

Proof. Start with a regular $(k, [n])$-scheme. Flip one coin to decide whether or not to reverse the meanings of 'true' and 'false'. Then flip a second coin to decide whether to reverse the order of the ground set and consider it in the order $\{n, n-1, \ldots, 1\}$. □

3. Uniform schemes and exact results

We are now ready to tackle the cases $k = 1$ and $k = 2$. The notion of a uniform scheme will enable us to present very simple arguments.

Theorem 3.1. $\lambda_1 = \frac{1}{4}$.

Proof. The coin-flip example discussed in Section 1 shows that $\lambda_1 \geqslant \frac{1}{4}$. We now show that $\lambda_1 \leqslant \frac{1}{4}$. In view of Theorem 2.4, it suffices to prove that, for every $\epsilon > 0$, there exists an integer n_0 so that if $n \geqslant n_0$ and $\mathscr{S}$ is any uniform $(1, [n])$-scheme, then $\lambda(\mathscr{S}) < \frac{1}{4} + \varepsilon$.

Let $\varepsilon > 0$. Then set n_0 as the least positive integer so that $n_0 \geqslant 2$ and $\frac{1}{4(n_0+1)} < \varepsilon$. Now let n be any integer with $n \geqslant n_0$ and let $\mathscr{S}$ be a uniform $(1, [n])$-scheme in an arbitrary probability space Ω.

Put $\mathbf{X} := \sum_{i=1}^{n} \mathbf{X}_i$ where $\mathbf{X}_i = 1$ if i is true and -1 otherwise. Then

$$0 \leqslant \mathrm{E}[\mathbf{X}^2] = \sum_{i=1}^{n} \mathrm{E}[\mathbf{X}_i^2] + \sum_{1 \leqslant i,j \leqslant n, i \neq j} E[\mathbf{X}_i \mathbf{X}_j]. \tag{3.1}$$

Now

$$\sum_{i=1}^{n} \mathrm{E}[\mathbf{X}_i^2] = \sum_{i=1}^{n} 1 = n.$$

Furthermore, for each $i, j = 1, 2, \ldots, n$ with $i \neq j$,

$$\mathrm{E}[\mathbf{X}_i\mathbf{X}_j] = \Pr[ij] - \Pr[i\bar{j}] - \Pr[\bar{i}j] + \Pr[\bar{i}\bar{j}].$$

Since $\mathscr{S}$ is uniform,

$$\Pr[i\bar{j}] = \Pr[\bar{i}j] = \lambda(\mathscr{S}).$$

Noting that

$$\Pr[ij] + \Pr[i\bar{j}] + \Pr[\bar{i}j] + \Pr[\bar{i}\bar{j}] = 1,$$

so that

$$\Pr[ij] + \Pr[\bar{i}\bar{j}] = 1 - 2\lambda(\mathscr{S}),$$

and

$$\mathrm{E}[\mathbf{X}_i\mathbf{X}_j] = 1 - 4\lambda(\mathscr{S}),$$

we can rewrite inequality (3.1) as

$$0 \leqslant n + n(n-1)(1 - 4\lambda(\mathscr{S})).$$

Thus,

$$4n(n-1)\lambda(\mathscr{S}) \leqslant n^2$$

and

$$\lambda(\mathscr{S}) \leqslant \frac{1}{4} + \frac{1}{4(n-1)} < \frac{1}{4} + \varepsilon. \qquad \square$$

Theorem 3.2. $\lambda_2 = \frac{1}{3}$.

Proof. We first show that $\lambda_2 \geqslant \frac{1}{3}$. Let $\mathscr{F}$ consist of all linear orders '$\prec$' on $[n]$. Choose one of these orders, say L, uniformly at random. Then define $\{i, j\} \in \binom{[n]}{2}$ to be true when $i \prec j$ in L, that is, when the L-order and the natural order agree for i and j. We then have that for all $(2, n)$-shift pairs (A, B), $\Pr[A\overline{B}] = \frac{1}{3}$. To see this, observe that if $1 \leqslant i < j < k \leqslant n$, $A = \{i, j\}$ and $B = \{j, k\}$, $\Pr[A\overline{B}]$ is just the probability that $i \prec j$ and $k \prec j$ in L, which is clearly $\frac{1}{3}$.

We now prove that $\lambda_2 \leqslant \frac{1}{3}$. Let $\varepsilon > 0$. In view of Theorem 2.4, it is enough to show that if n is sufficiently large, and $\mathscr{S}$ is any uniform $(2, [n])$-scheme, then $\lambda(\mathscr{S}) < \frac{1}{3} + \varepsilon$.

So suppose that $\mathscr{S}$ is a uniform $(2, [n])$ scheme in a probability space Ω.

Set $q = \Pr[\{1,2\}\overline{\{2,3\}}] = \Pr[\overline{\{1,2\}}\{2,3\}]$. Our goal is then to show that $q < \frac{1}{3} + \varepsilon$. Suppose to the contrary that $q \geqslant \frac{1}{3} + \varepsilon$. We argue to a contradiction, provided of course that n is sufficiently large.

There are 8 $(2, 3)$ patterns in this scheme. Since $\mathscr{S}$ is uniform, we conclude that

$$\begin{aligned}
\frac{q}{2} &= \Pr[\{1,2\}\{1,3\}\overline{\{2,3\}}] \\
&= \Pr[\{1,2\}\overline{\{1,3\}\{2,3\}}] \\
&= \Pr[\overline{\{1,2\}}\{1,3\}\{2,3\}] \\
&= \Pr[\overline{\{1,2\}\{1,3\}}\{2,3\}].
\end{aligned}$$

Let q_1 and q_2 satisfy

$$\begin{aligned} q_1 &= \Pr[\{1,2\}\{1,3\}\{2,3\}] \\ &= \Pr[\overline{\{1,2\}\{1,3\}\{2,3\}}], \\ q_2 &= \Pr[\{1,2\}\overline{\{1,3\}}\{2,3\}] \\ &= \Pr[\overline{\{1,2\}}\{1,3\}\overline{\{2,3\}}]. \end{aligned}$$

Let $m = \lfloor (n-1)/2 \rfloor$ and, for each $i \in [m]$, define random variables $\mathbf{X}_i$ and $\mathbf{Y}_i$ as follows. Set $\mathbf{X}_i = 1$ when $\{i, m+1\}$ is true, and 0 otherwise; set $\mathbf{Y}_i = 1$ when $\{m+1, m+1+i\}$ is *false*, and 0 otherwise. Then put $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_m$ and $\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_m$. Since the expectation of $(\mathbf{X} - \mathbf{Y})^2$ is nonnegative, we have $\mathrm{E}[\mathbf{X}^2] + \mathrm{E}[\mathbf{Y}^2] \geqslant 2\mathrm{E}[\mathbf{XY}]$. It follows that

$$m + m(m-1)(\Pr[\{1,3\}\{2,3\}] + \Pr[\overline{\{1,2\}\{1,3\}}]) \geqslant 2m^2 q. \tag{3.2}$$

Now

$$\Pr[\{1,3\}\{2,3\}] = \Pr[\overline{\{1,2\}}\{1,3\}\{2,3\}] + \Pr[\{1,2\}\{1,3\}\{2,3\}] = \frac{q}{2} + q_1.$$

Similarly,

$$\Pr[\overline{\{1,2\}\{1,3\}}] = \Pr[\overline{\{1,2\}\{1,3\}}\{2,3\}] + \Pr[\overline{\{1,2\}\{1,3\}\{2,3\}}] = \frac{q}{2} + q_1.$$

Thus inequality (3.2) becomes

$$m + m(m-1)(q + 2q_1) \geqslant 2m^2 q. \tag{3.3}$$

We also know $4\frac{q}{2} + 2q_1 + 2q_2 = 1$. Solving for q_1 and substituting into inequality (3.3), we obtain

$$m^2 \geqslant m + m(m-1)(1 - 2q_2) \geqslant (3m^2 - m)q. \tag{3.4}$$

This implies that $q \leqslant \frac{1}{3} + \frac{1}{9m-3}$, which completes the proof. □

The reader may have noticed that we have only proved that $\lambda(2,n) \geqslant \frac{1}{3}$, but in fact $\lambda(2,n) > \frac{1}{3}$. To see this, condition on the linear order L not being the natural order.

4. Uniform schemes and lower bounds

In this section, we generalize the constructions for λ_1 and λ_2 to show that $\lambda_k \geqslant \frac{1}{2} - \frac{1}{2k+2}$, for all $k \geqslant 1$. Thus $\lambda_3 \geqslant \frac{3}{8}$ and $\lambda_4 \geqslant \frac{2}{5}$, and we believe these two inequalities are tight. However, we also show that $\lambda_5 \geqslant \frac{27}{64} > \frac{5}{12}$, so the obvious conjecture fails.

Theorem 4.1. *$\lambda_k \geqslant \frac{1}{2} - \frac{1}{2k+2}$, for all $k \geqslant 1$.*

Proof. The proposed lower bound holds when $k = 1$ and $k = 2$ by our previous remarks, so we may assume $k \geqslant 3$. For each $n \geqslant k+1$, we construct a uniform $(k,[n])$-scheme $\mathscr{S}$ with $\lambda(\mathscr{S}) = \frac{1}{2} - \frac{1}{2k+2}$ as follows. To each $i \in [n]$, we assign independently a random bit b_i and a random real r_i, drawn (say) from the uniform distribution on $[0,1]$. Let $A = \{i_1, \ldots, i_k\}$ be an event in the scheme, and define $j = j(A)$ as the index for which r_{i_j}

is maximized. We then say that A is true if either $b_{i_j} = 1$ and j is odd, or $b_{i_j} = 0$ and j is even.

Now let (A, B) be a $(k, [n])$-shift pair, and let $A \cup B = \{i_1, i_2, \ldots, i_{k+1}\}$. Then the probability that the largest real in $\{r_{i_j} : 1 \leqslant j \leqslant k+1\}$ belongs to $\{r_i : 2 \leqslant j \leqslant k\}$ is $\frac{k-1}{k+1}$, and when this occurs, the probability that A is true and B is false is $\frac{1}{2}$. On the other hand, the probability that the largest real in $\{r_{i_j} : 1 \leqslant j \leqslant k+1\}$ belongs to $\{r_{i_1}, r_{i_{k+1}}\}$ is $\frac{2}{k+1}$ and in this case, the probability that A is true and B is false is $\frac{1}{4}$. It follows that $\lambda(\mathscr{S}) = \frac{k-1}{k+1} \cdot \frac{1}{2} + \frac{2}{k+1} \cdot \frac{1}{4} = \frac{1}{2} - \frac{1}{2k+2}$. □

Again, we note that we can improve the argument given in the preceding theorem to show that $\lambda(k, n) > \frac{1}{2} - \frac{1}{2k+2}$, although the difference goes to zero as n increases. However, we were somewhat surprised to find that for $k = 5$ we could beat $\frac{1}{2} - \frac{1}{2k+2}$ by a constant.

Theorem 4.2. $\lambda_5 \geqslant \frac{27}{64}$.

Proof. For this result, we need only a random bit b_i for each $i \in [n]$, thus a vector $v(A) \in \{0, 1\}^5$ for each event A. Suppose we take A to be true just when $h(v(A), (10101)) \leqslant 2$, where

$$h(u, v) := |\{i : u_i \neq v_i\}|$$

is the Hamming distance for binary sequences. Then it is easily checked that $\Pr[A\overline{B}] = \frac{13}{32}$, slightly below the $\frac{5}{12}$ lower bound of Theorem 4.1. However, we can improve the calculation slightly by interchanging the roles of two of the vectors, namely 10011 and 00110. In other words, A is now false when $v(A) = 10011$ and true when $v(A) = 00110$, but otherwise the Hamming rule still applies.

The result is that the number of 6-bit vectors which give $A\overline{B}$ for a shift pair (A, B) goes up from 26 to 27, giving $\lambda_5 \geqslant \frac{27}{64} > \frac{5}{12}$. □

We now show that λ_k is a strictly increasing function of k.

Proposition 4.3. *For every* $k \geqslant 1$, $\lambda_{k+1} > \lambda_k$.

Proof. Let n and k be integers, with $n \geqslant k+2$, let Ω be a probability space and let $\mathscr{S}$ be a uniform $(k, [n])$-scheme. We construct a probability space Ω' and a uniform $(k+1, [n])$-scheme $\mathscr{S}'$ in Ω' with

$$\lambda(\mathscr{S}') > \lambda(\mathscr{S}) + f(k),$$

where $f(k)$ is a positive quantity depending only on k.

Choose independent samples $\mathbf{X}_1, \mathbf{X}_2, \ldots$ from Ω. For any $(k, [n])$-shift pair (A, B) in Ω, the probability space Ω' is to contain an event $A' = A \cup B$. In Ω', we define the probability of A' as follows. Let $t = t(A \cup B)$ be least integer such that the events A and B differ in truth value in $\mathbf{X}_t$, and let A' be true in Ω' just when A is true in $\mathbf{X}_t$. Evidently, the resulting $(k+1, [n])$-scheme $\mathscr{S}'$ is regular.

Now let (A', B') be a $(k+1, [n])$-shift pair. Note that (A', B') corresponds to a two $(k, [n])$-shift pairs, which we denote (A, B) and (B, C), so that $A' = A \cup B$ and $B' = B \cup C$.

Since $\mathscr{S}$ is regular, we may choose numbers q_1, q_2 and q_3, with $q_1 + q_2 + q_3 = 1$ so that in the probability space Ω:

$$\begin{aligned} q_1 &= \Pr[ABC] = \Pr[\overline{ABC}], \\ q_2 &= \Pr[A\overline{B}C] = \Pr[\overline{A}B\overline{C}], \quad \text{and} \\ q_3 &= \Pr[AB\overline{C}] = \Pr[\overline{A}BC] = \Pr[\overline{AB}C] = \Pr[A\overline{BC}]. \end{aligned}$$

Therefore, in Ω,

$$\lambda(\mathscr{S}) = \Pr[A\overline{B}] = \Pr[A\overline{B}C] + \Pr[A\overline{BC}] = q_2 + q_3.$$

On the other hand, in Ω',

$$\begin{aligned} \lambda(\mathscr{S}') &= \Pr[A'\overline{B'}] \\ &= \frac{\left(\frac{1}{2} \cdot \Pr[A\overline{BC}] + \frac{1}{2} \cdot \Pr[\overline{AB}C] + \Pr[A\overline{B}C]\right)}{1 - 2q_1} \\ &= \frac{q_2 + q_3}{1 - 2q_1}. \end{aligned}$$

To complete the proof, we analyse the quantity $2q_1$ and show that it is positive by an amount depending only on k and not on n.

Let m be the least positive integer (guaranteed by Ramsey's theorem) so that, if the k-element subsets of an m-element set S are coloured with two colours, say T and F, then there is a $k+2$ element subset $H \subseteq S$ so that all k-element subsets of H receive the same colour.

Now a sampling from Ω produces a colouring of the k-element subsets of $[n]$. Given a sampling, say that a subset $H \subset [n]$ is *homogeneous* if either (1) all of the k-element subsets of H are true, or (2) all of the k-element subsets of H are false. Note that $2q_1 = \Pr[ABC] + \Pr[\overline{ABC}]$, so our requirement that *all* k-element subsets of H have the same truth value is stronger than what is required.

There are $\binom{n}{m}$ subsets of size m, and each of these subsets in turn contains a k-element homogeneous subset H. Any $k+2$-element subset of $[n]$ is a subset of exactly $\binom{n-k+2}{m-k-2}$ different m-element subsets of $[n]$. Since

$$\binom{n}{k+2}\binom{n-k-2}{m-k-2} = \binom{n}{m}\binom{m}{k-2},$$

it follows that, for every $k+2$-element subset $T \subset [n]$, the probability that T is homogeneous in a sampling from Ω is at least $\binom{m}{k-2}^{-1}$, which is independent of n as required. $\square$

5. Shift graphs and upper bounds

When $1 \leqslant k < n$, we define the (k,n)-*shift graph* $\mathbf{S}(k,n)$ as the graph whose vertex set is $\binom{[n]}{k}$ with a k-element set A adjacent to a k-element set B in $\mathbf{S}(k,n)$ exactly when (A,B) is a $(k,[n])$-shift pair. Note that $\mathbf{S}(1,n)$ is a complete graph on n vertices, but for $k \geqslant 2$, $\mathbf{S}(k,n)$ is triangle-free. Historically, the graphs $\mathbf{S}(2,n)$ have been called *shift graphs*, and $\mathbf{S}(3,n)$ *double shift graphs*.

We now provide an upper bound for λ_k which, together with our lower bound of Theorem 4.1, shows that $\frac{1}{2} - \lambda_k = \Theta(\frac{1}{k})$.

Theorem 5.1. $\lambda_k \leqslant \frac{1}{2} - \frac{1}{4k+2}$, *for all* $k \geqslant 1$.

Proof. Let us note first that the shift graph $\mathbf{S}(k, 2k+1)$ contains a cycle of length $2k+1$. To see this, note the obvious path of length $k+1$,

$$\{1, \ldots, k\}, \{2, \ldots, k+1\}, \ldots, \{k+2, \ldots, 2k+1\},$$

in the k-shift graph G on vertex set $\{1, \ldots, 2k+1\}$. The subgraph of G induced on vertices $\{1, \ldots, k\} \cup \{k+2, \ldots, 2k+1\}$ is of course isomorphic to $\mathbf{S}(k, 2k)$ and contains a path of length k with the same endpoints. Linking these two paths produces the desired cycle.

Suppose now that Ω is a probability space and that $\mathscr{S}$ is a uniform $(k, [n])$-scheme in Ω, with $n \geqslant 2k+1$. Assuming $n \geqslant 2k+1$, we can concentrate on events corresponding to k-element subsets which form a cycle of length $2k+1$ in the shift graph $\mathbf{S}(k, n)$. Observe that in a sampling from Ω, only $2k$ of the $2k+1$ consecutive sets (vertices) on this cycle can have differing truth values. It follows that $2\lambda(\mathscr{S}) \leqslant \frac{2k}{2k+1}$, so that $\lambda(\mathscr{S}) \leqslant \frac{1}{2} - \frac{1}{4k+2}$, as required. □

Using the technique employed in the proof of Theorem 3.2, it is straightforward to improve the upper bound on λ_k to $\frac{1}{2} - \frac{1}{4k-2}$ when $k \geqslant 2$. We leave this as an exercise.

6. Shift graphs and dimension theory

In the next two sections of this paper, we present a brief discussion of the combinatorial problems that motivated this line of research. We begin with a review of the chromatic number of shift graphs. Historically, the shift graphs $\mathbf{S}(2, n)$ were an important instance of triangle-free graphs with large chromatic number.

Note that it follows immediately from Ramsey's theorem that, for every $k \geqslant 1$ and every r, there exists n_0 so that the chromatic number of $\mathbf{S}(k, n) \geqslant r$, when $n \geqslant n_0$. For $k = 1$, this statement is trivial, since the shift graph $\mathbf{S}(1, n)$ is a complete graph, but for $k \geqslant 2$, it is a bit more surprising.

The formula for the chromatic number of $\mathbf{S}(2, n)$ is now considered to be folklore, although it has frequently been attributed to Andras Hajnal. (We use the notation '$\lg n$' as shorthand for $\log_2 n$.)

Proposition 6.1. *The chromatic number* $\chi(\mathbf{S}(2, n))$ *of the shift graph* $\mathbf{S}(2, n)$ *is exactly* $\lceil \lg n \rceil$.

For double shift graphs, we have the following estimate.

Proposition 6.2. *The chromatic number* $\chi(\mathbf{S}(3, n))$ *of the double shift graph satisfies*

$$\chi(\mathbf{S}(3, n)) = \lg \lg n + \left(\frac{1}{2} + o(1)\right) \lg \lg \lg n. \tag{6.1}$$

□

When $\mathbf{P} = (X, P)$ is a poset, a linear order L on X is called a *linear extension* of P when $x < y$ in L for all $x, y \in X$ with $x < y$ in P. A set $\mathscr{R}$ of linear extensions of P is called a *realizer* of $\mathbf{P}$ when $P = \cap\mathscr{R}$, that is, for all x, y in X, $x < y$ in P if and only if $x < y$ in L for every $L \in \mathscr{R}$. The minimum cardinality of a realizer of $\mathbf{P}$ is called the *dimension* of $\mathbf{P}$ and is denoted dim($\mathbf{P}$). We refer the reader to the monograph [15] for additional background material on dimension theory,

A poset $\mathbf{P} = (X, P)$ is called an *interval order* if there exists a family $\{[a_x, b_x] : x \in X\}$ of nonempty closed intervals of $\mathbb{R}$ so that $x < y$ in P if and only if $b_x < a_y$ in $\mathbb{R}$. The interval order $\mathbf{I}_n$ consisting of all intervals with integer endpoints from $[n]$ is called the *canonical interval order*.

Although general posets of height 2 can have arbitrarily large dimension, for interval orders, large height is required for large dimension. The following result is due to Füredi, Hajnal, Rödl and Trotter [10].

Theorem 6.3. *If* $\mathbf{P} = (X, P)$ *is an interval order of height n, then*

$$\dim(\mathbf{P}) \leqslant \lg\lg n + (1/2 + o(1)) \lg\lg\lg n. \tag{6.2}$$

□

As noted in [10], the family of canonical interval orders witnesses that the inequality in the preceding theorem is best possible.

7. Fractional dimension and Ramsey theory for probability spaces

In many instances, it is useful to consider a fractional version of an integer valued combinatorial parameter, as in many cases, the resulting *LP* relaxation sheds light on the original problem. In [2], Brightwell and Scheinerman proposed to investigate fractional dimension for posets.

Let $\mathbf{P} = (X, P)$ be a poset and let $\mathscr{F} = \{M_1, \ldots, M_t\}$ be a multiset of linear extensions of P. Brightwell and Scheinerman [2] call $\mathscr{F}$ a *k-fold realizer* of P if, for each incomparable pair (x, y), there are at least k linear extensions in $\mathscr{F}$ which reverse the pair (x, y), that is, $|\{i : 1 \leqslant i \leqslant t,\ x > y \text{ in } M_i\}| \geqslant k$. The *fractional dimension* of $\mathbf{P}$, denoted by fdim($\mathbf{P}$), is then defined as the least real number $q \geqslant 1$ for which there exists a k-fold realizer $\mathscr{F} = \{M_1, \ldots, M_t\}$ of P so that $k/t \geqslant 1/q$ (it is easily verified that the least upper bound of such real numbers q is indeed attained and is therefore a rational number). Using this terminology, the *dimension* of $\mathbf{P}$ is just the least t for which there exists a 1-fold realizer of P. It follows immediately that fdim($\mathbf{P}$) $\leqslant$ dim($\mathbf{P}$), for every poset $\mathbf{P}$.

The dimension or fractional dimension of a *class* of posets is defined to be the least upper bound of dim($\mathbf{P}$) (respectively, fdim($\mathbf{P}$)) over all posets $\mathbf{P}$ in the class. We have seen that $\dim(\mathscr{I}) = \infty$ for the class $\mathscr{I}$ of interval orders, but Brightwell and Scheinerman showed that $\text{fdim}(\mathscr{I}) \leqslant 4$. To see this, observe that if $\mathbf{P} = (X, P)$ is an interval order and $A \subset X$, there is a linear extension L of P with $x > y$ in L for any incomparable pair (x, y) with $x \in A$ and $y \notin A$. Building a realizer from one such L for each subset A of X of size $\lfloor |X|/2 \rfloor$ gives fdim($\mathbf{P}$) < 4.

Brightwell and Scheinerman conjectured in [2] that $\mathrm{fdim}(\mathscr{I}) = 4$, even though no example of an interval order of fractional dimension even as high as 3 was then known. Using the techniques developed in the preceding sections, we can now settle this conjecture in the affirmative.

Theorem 7.1. $\mathrm{fdim}(\mathscr{I}) = 4$.

Proof. Let $0 < \varepsilon < \frac{1}{2}$. We show that, for large n, the fractional dimension of the canonical interval order $\mathbf{I}_n$ consisting of all intervals with integer end points from $[n]$ satisfies $4 - \varepsilon < \mathrm{fdim}(\mathbf{I}_n)$. To the contrary, suppose that $\mathrm{fdim}(\mathbf{I}_n) \leqslant 4 - \varepsilon$, regardless of the size of n. We argue to a contradiction, provided n is sufficiently large.

We begin by using Theorem 3.1 ($\lambda_1 = \frac{1}{4}$) to obtain an m_0 such that, whenever $\{U_i : 1 \leqslant i \leqslant m_0\}$ is a sequence of events, there exist integers i, j with $1 \leqslant i < j \leqslant m_0$ with $\Pr[U_i\overline{U_j}] < \frac{1}{4} + \frac{\varepsilon}{32}$.

Let $\mathscr{F}$ be a multiset of linear extensions of $\mathbf{I}_n$ which witnesses that $\mathrm{fdim}(\mathbf{I}_n) \leqslant 4 - \varepsilon$. Let Ψ be the probability space given by the uniform distribution on $\mathscr{F}$, so that the probability of an event $E \subset \mathscr{F}$ is just $|E|/|\mathscr{F}|$. For intervals $A = [a_1, a_2]$ and $B = [b_1, b_2]$, we then let $A > B$ denote the event consisting of those linear extensions $L \in \mathscr{F}$ for which $A > B$ in L. When $a_2 < b_1$, note that $\Pr[A > B] = 0$; similarly, when $b_2 < a_1$, then $\Pr[A > B] = 1$. Otherwise, $\frac{1}{4} + \frac{\varepsilon}{16} < \frac{1}{4-\varepsilon} < \frac{1}{4}\Pr[A > B] < 1 - \frac{1}{4-\varepsilon}$. In what follows, we will concentrate on the implications that follow from the inequality $\Pr[A > B] > \frac{1}{4} + \frac{\varepsilon}{16}$ when $a_1 < b_1 < a_2 < b_2$.

Next, let $\delta = \frac{\varepsilon}{128}$, and let $m_1 = 1 + \lfloor \frac{1}{\delta} \rfloor$. We use the (by now) standard trick of approximating probabilities in Ψ with one of the m_1 discrete values from $Q = \{i\delta : 0 \leqslant i < m_1\}$. For an event E, we let $\mathrm{P}[E] = i\delta$, where i is the largest integer for which $i\delta \leqslant \Pr[E]$. Of course, $\mathrm{P}[E] \leqslant \Pr[E] < \mathrm{P}[E] + \delta$, for every event $E \in \Psi$.

Now let r be a positive integer, and let $W = \{w_1, w_2, \ldots, w_{4r+2}\} \in \binom{[n]}{4r+2}$. Then define an interval $A(W) = [a_1, a_2]$ with $a_1 = w_{r+1}$ and $a_2 = w_{3r+2}$. Also define subsets $S(W)$, $T(W)$, $U(W)$, $V(W)$ from $\binom{[n]}{r}$ by setting

$$\begin{aligned} S(W) &= \{w_1, w_2, \ldots, w_r\}, \\ T(W) &= \{w_{r+2}, w_{r+3}, \ldots, w_{2r+1}\}, \\ U(W) &= \{w_{2r+2}, w_{2r+3}, \ldots, w_{3r+1}\}, \quad \text{and} \\ V(W) &= \{w_{3r+3}, w_{3r+4}, \ldots, w_{4r+2}\}. \end{aligned}$$

Then let $\mathrm{Down}(W)$ denote the event consisting of all linear extensions $L \in \mathscr{F}$ for which there exist integers $s \in S(W)$, $t \in T(W)$ with $[s,t] > [a_1, a_2]$ in L. Also, let $\mathrm{Up}(W)$ denote the event consisting of all $L \in \mathscr{F}$ for which there exist integers $u \in U(W)$, $v \in V(W)$ with $[a_1, a_2] > [u, v]$ in L. Evidently, $\mathrm{Down}(W) \cap \mathrm{Up}(W) = \emptyset$.

For a fixed value of r, our Ramsey-theoretic techniques would enable us to say that we may assume that there exist integers i and j so that for every set $W \in \binom{[n]}{4r+2}$, $\mathrm{P}[\mathrm{Down}(W)] = i\delta$ and $\mathrm{P}[\mathrm{Up}(W)] = j\delta$. In fact, we can make this assumption for a large (but bounded) set of distinct values of r. However, we can make much stronger assumptions; in particular, we will argue that (up to a small error) the events $\mathrm{Down}(W)$

and Up(W) depend *only* on $A(W)$, that is, the events Down(W) and Up(W) are (almost entirely) independent of the integer r and the elements of $W - A(W)$. To make this last statement more precise, we require some additional notation.

Set $m_2 = 2m_1$ and $m_3 = (2m_0 + 1)(m_2 + 1) - 1$. It follows from Ramsey's theorem that if n is sufficiently large, there exist values $d_1, \ldots, d_{m_2}, u_1, u_2, \ldots, u_{m_2}$ from the m_1-element set Q and a subset $H \in \binom{[n]}{m_3}$ so that for every $r = 1, 2, \ldots, m_2$ and every set $W \in \binom{H}{4r+2}$, P[Down(W)] $= d_r$ and P[Up(W)] $= u_r$.

Claim 1. $d_{r-1} \leqslant d_r$, for $r = 2, 3, \ldots, m_2$.

To verify this claim, consider sets $W_1 \in \binom{H}{4r+2}$ and $W_2 \in \binom{H}{4r-2}$ where W_1 is the subset of W_2 obtained by deleting the largest elements in the four $r+1$-element subsets $S(W_2)$, $T(W_2)$, $U(W_2)$, and $V(W_2)$ of W_2. In this case, we observe that Down(W_1) $\subseteq$ Down(W_2).

Dually, we can also verify the following claim.

Claim 2. $u_{r-1} \leqslant u_r$, for $r = 2, 3, \ldots, m_2$.

The next result follows easily from the fact that the sequences $\{d_i\}$ and $\{u_i\}$ can increase at most $m_1 - 1$ times.

Claim 3. There is an integer r with $1 < r \leqslant m_2$ and elements $d, u \in Q$ so that $d_{r-1} = d_r = d$ and $u_{r-1} = u_r = u$.

Next, we relabel our homogeneous set H as $\{1, 2, \ldots, m_3\}$ and consider the subset $K = \{j(m_2 + 1) : 1 \leqslant j \leqslant 2m_0\} = \{k_1 < k_2 < \cdots < k_{2m_0}\}$. The important fact about the subset K is that, between any two elements of K, there are m_2 elements of $H - K$. In addition, there are m_2 elements of $H - K$ occurring before the least element of K, and m_2 elements of $H - K$ after the largest element of K.

Now let i be any integer with $1 \leqslant i \leqslant m_0$. Define a subset $W_i \in \binom{H}{4r-2}$ by first taking $A(W_i) = [k_i, k_{m_0+i}]$. Then define $S(W_i) = \{i(m_2 + 1) - j : 1 \leqslant j \leqslant r - 1\}$, $T(W_i) = \{i(m_2 + 1) + j : 1 \leqslant j \leqslant r - 1\}$, $U(W_i) = \{(m_0 + i)(m_2 + 1) - j : 1 \leqslant j \leqslant r - 1\}$, and $V(W_i) = \{(m_0 + i)(m_2 + 1) + j : 1 \leqslant j \leqslant r - 1\}$. Note that the end points of $A(W_i)$ belong to K, but $W_i - A(W_i) \subset H - K$.

Now let j be any integer with $i < j \leqslant m_0$, and consider the event E_1 consisting of all linear extensions $L \in \mathscr{F}$ for which $A(W_i) > A(W_j)$ in L but $L \notin$ Up(W_i).

Claim 4. $\Pr[E_1] < 2\delta$.

To verify this claim, we consider a second set W_i' formed from W_i by adding k_j to $U(W_i)$ and k_{m_0+j} to $V(W_i)$. Also, choose the largest element s_0 of $H - W_i$ satisfying $s_0 < S$ and add s_0 to $S(W_i)$. Dually, let t_0 be the least element of $H - W_i$ satisfying $t_0 > T$ and add t_0 to T. We know that P[Up(W_i)] $=$ P[Up(W_i')]. Also, we know that both Up(W_i) and $A > B$ are subsets of Up(W_i'). So the claim follows.

Dually, let E_2 be the event consisting of all linear extensions in $\mathscr{F}$ for which $A(W_i) > A(W_j)$ in L and $L \notin \text{Down}(W_j)$. Then the argument given above also establishes the following claim.

Claim 5. $\Pr[E_2] < 2\delta$.

Combining the previous claims, we obtain the following inequality.

Claim 6. For all integers i, j with $1 \leqslant i < j \leqslant m_0$, $\Pr[A(W_i) > A(W_j)] < \Pr[\text{Up}(W_i)\overline{\text{Up}(W_j)}] + 4\delta$.

We are now ready to obtain the final contradiction. We know that there is some pair i, j with $1 \leqslant i < j \leqslant m_0$ for which $\Pr[\text{Up}(W_i)\overline{\text{Up}(W_j)}] < \frac{1}{4} + \frac{\varepsilon}{32}$. However, $4\delta < \frac{\varepsilon}{32}$, which then implies that $\Pr[A(W_i) > A(W_j)] < \frac{1}{4} + \frac{\varepsilon}{16}$. The contradiction completes the proof. □

In a certain sense, the preceding theorem signals an interesting departure from the analogy between interval orders and double shift graphs. In retrospect, the dimension problem for interval orders is more or less the same as the chromatic number problem for double shift graphs. It took some 15 years to see clearly how to make the translation, but this is by now completely understood.

In contrast, we have now seen that while the *fractional* chromatic number of the double shift graph is at most $\frac{8}{3}$, the fractional dimension of an interval order can be arbitrarily close to 4.

8. Random variables

We now return to the $k = 1$ case, where events are labelled by numbers $1, \ldots, n$, with the idea of weakening the hypothesis and strengthening the conclusion of Theorem 3.1. Suppose that the coin-flip example is modified as follows: first a coin is chosen at random from a box of coins, some of which are badly bent and thus biased. Then the chosen coin is flipped n times as before. It is then immediate that $\Pr[i\bar{j}] \leqslant \frac{1}{4}$ for each $i < j$ since whatever coin is chosen, the conditional probability of $i\bar{j}$ is $p(1-p) \leqslant \frac{1}{4}$ where p is the probability of heads.

If we set $X_i = 1$ when the ith toss is 'heads' and 0 otherwise, the sequence $X_1, \ldots, X_n$ forms what is called an 'i.i.d. mix' – the result of an experiment in which independent samples are drawn from a distribution chosen once from some distribution on distributions. Theorem 3.1 would then follow, for example, from the statement that any long list of Bernoulli random variables contains a pair which are nearly an i.i.d. mix; in fact, we can prove more.

Any i.i.d. mix $X_1, \ldots, X_n$ is also *exchangeable*, meaning that $X_{\pi(1)}, \ldots, X_{\pi(n)}$ has the same joint distribution as $X_1, \ldots, X_n$ for any permutation $\pi \in S_n$. Exchangeability does *not* imply i.i.d. mix; for example, in the case where we flip a fair coin and condition on exactly $n/2$ heads, we retain exchangeability but force an event ('$\sum X_i = n/2$') which has low probability in any i.i.d. mix.

However, any reasonably well-behaved infinite sequence of random variables, every initial segment of which is exchangeable, *is* equivalent to an infinite i.i.d. mix; this is the celebrated theorem of de Finetti [8]. Finite versions do exist but are necessarily weakened; the following is found in Diaconis and Freeman [5].

Theorem 8.1. *Let* $X_1, \ldots, X_k, \ldots, X_r$ *be an exchangeable sequence of random variables with values in the set* $[m] = \{1, 2, \ldots, m\}$. *Then there is an i.i.d. mix of length* k *at variation distance at most* $2mk/r$ *from* $X_1, \ldots, X_k$.

The bound $2mk/r$ is in fact best possible and represents the variation distance between sampling with and without replacement; in our example, with and without conditioning on exactly $r/2$ heads out of r coin-flips. Theorem 8.1 applies of course to *any* length-k subsequence of $X_1, \ldots, X_r$, whereas we seek only assertions about *some* subsequence, but we must get rid of the exchangeability assumption.

An infinite sequence $\{X_i\}$ of random variables is said to have the *selection* property (see, for instance, [11]) if, for any k and any $1 \leqslant j_1 < \cdots < j_k$, the joint distribution of $X_{j_1}, \ldots, X_{j_k}$ is the same as for $X_1, \ldots, X_k$. For infinite sequences, selection implies exchangeability and thus serves as an alternate hypothesis for de Finetti's theorem.

Ramsey's theorem provides a finite, approximate version of selection.

Theorem 8.2. *Fix integers* m *and* $k \leqslant r$, *and let* $\varepsilon > 0$. *Then for* n *sufficiently large, every sequence* $X_1, \ldots, X_n$ *of random variables with values in* $[m]$ *contains a subsequence* $X_1^*, \ldots, X_r^*$ *such that any two subsequences of* X_i^* *of length* k *have the same distribution, up to variation distance* ε.

We now need to convert the selection property to exchangeability. Two Bernoulli random variables with the same distribution are exchangeable, but things get more complex already with 3-valued r.v.s: if (X, Y) takes on values (0,1), (1,2) and (2,0) with equal probability, then X and Y have the same (uniform) distribution but are not even approximately exchangeable.

Nonetheless, approximate exchangeability is achievable for us because, for any fixed k and m, if every k-subsequence of a large number of $[m]$-valued random variables has (or nearly has) the same distribution, then that distribution must be *approximately symmetric*; that is, for any sequence $a_1, \ldots, a_k$ of range values and permutation π of the range, $\Pr[\bigwedge_i (X_i = a_i)] \sim \Pr[\bigwedge_i (X_i = a_{\pi(i)})]$.

The '$k = 2$' version of this fact is proved by Komlós [14] and employed to prove the following theorem.

Theorem 8.3. *For every positive integer* m *and* $\varepsilon > 0$ *there is an* n *such that if* $X_1, \ldots, X_n$ *are random variables with values in* $[m]$, *then there are indices* i, j *with* $1 \leqslant i < j \leqslant n$ *such that, for any* $a, b \in M$,

$$|\Pr[(X_i = a) \wedge (X_j = b)] - \Pr[(X_i = b) \wedge (X_j = a)]| < \varepsilon.$$

Our theorem is the result of putting together Theorem 8.2 with an approximate version of Theorem 8.1 and a generalization of Theorem 8.3.

Theorem 8.4. *For every m and k and every $\varepsilon > 0$ there is an n such that if $X_1, \ldots, X_n$ are random variables with values in $[m]$, then there is a subsequence $X_1^*, \ldots, X_k^*$ of the X_is at variation distance at most ε from some i.i.d. mix.*

Proof. The general strategy of the proof is as follows. We first apply Theorem 8.1 to choose r so that *any* k-subsequence of a *nearly exchangeable* list of r $[m]$-valued random variables will satisfy the conclusion of the theorem. We then apply Ramsey's theorem to get a very long subsequence of the X_is with an approximately common r-wise joint distribution. Finally we show that this distribution is approximately symmetric, therefore any length r sub-subsequence will be nearly exchangeable. We make no attempt to optimize the value of $n = n(m, k, \varepsilon)$.

Accordingly, using Theorem 8.1, we begin by choosing r large enough so that in any exchangeable sequence $Y_1, \ldots, Y_k, \ldots, Y_r$ of $[m]$-valued r.v.s the initial subsequence $Y_1, \ldots, Y_k$ lies within variation distance $\varepsilon/2$ of some i.i.d. mix. Choose

$$\delta = \left(\frac{\varepsilon}{22rm^r}\right)^2$$

and let $s = \lceil 1/\delta \rceil$. Being generous, we then have $\frac{1}{\delta} \leqslant s \leqslant \frac{3}{\delta}$. Employing Ramsey's theorem, let n be large enough so that, for every sequence $X_1, \ldots, X_n$ of $[m]$-valued r.v.s, there is a k-ary distribution $\mathbf{p} = \{p_{\hat{a}} : \hat{a} \in [m]^r\}$ and a subsequence $X_1^*, \ldots, X_{rs}^*$ such that, for every $1 \leqslant j_1 \leqslant \cdots \leqslant j_r \leqslant rs$ and every sequence $\hat{a} = (a_1, \ldots, a_r) \in [m]^r$,

$$\left| \Pr\left[\bigwedge_{1 \leqslant i \leqslant r} (X_{j_i}^* = a_i) \right] - p_{\hat{a}} \right| < \delta.$$

To show that $\mathbf{p}$ is almost symmetric, we need to break the X_i^*s into r blocks of length s and to introduce some new random variables. Set $I_i(a) = 1$ if $X_i^* = a$ and 0 otherwise, and for each j with $1 \leqslant j \leqslant r$ let

$$N_j(a) = \sum_{i=js+1}^{(j+1)s} I_i(a)$$

so that $N_j(a)$ counts the number of occurrences of the value a in the jth block.

Fixing $a \in [m]$ and $1 \leqslant j \leqslant r$, we have

$$\begin{aligned}
\mathrm{E}\left((N_j(a))^2\right) &= \mathrm{E}\left(\left(\sum_{i=js+1}^{(j+1)s} I_i(a)\right)^2\right) \\
&= \sum_{i=js+1}^{(j+1)s} \mathrm{E}\left((I_i(a))\right)^2 + 2 \sum_{js+1 \leqslant i < i' \leqslant (s+1)j} \mathrm{E}\left(I_i(a) I_{i'}(a)\right) \\
&\leqslant s + 2\binom{s}{2}(p_{aa} + \delta).
\end{aligned}$$

Thus, when $1 \leqslant j \leqslant j' \leqslant r$,

$$\begin{aligned}
\mathrm{E}\left((N_j(a) - N_{j'}(a))^2\right) &= \mathrm{E}\left((N_j(a))^2 + (N_{j'}(a))^2 - 2N_j(a)N_{j'}(a)\right) \\
&\leqslant 2\left(s + 2\binom{s}{2}(p_{aa} + \delta)\right) - 2\mathrm{E}\left(\sum_{i=js+1}^{(j+1)s} \sum_{i'=j's+1}^{(j'+1)s} I_i(a)I_{i'}(a)\right) \\
&\leqslant 2s + 2s(s-1)(p_{aa} + \delta) - 2s^2((p_{aa} - \delta) \\
&= 2s - 2sp_{aa} + 4s^2\delta - 2s\delta \\
&\leqslant 2s + 4s^2\delta \\
&\leqslant 8s.
\end{aligned}$$

Now we fix an arbitrary sequence $\hat{a} = (a_1, \ldots, a_r) \in [m]^r$ and let $\hat{b}$ differ from $\hat{a}$ by a single transposition; to save subscripts we assume the transposition involves the first two coordinates, setting $c = a_1 = b_2$ and $d = a_2 = b_1$ with $a_j = b_j$ for $j > 2$. We would like to show that

$$|p_{\hat{a}} - p_{\hat{b}}| \leqslant 9\sqrt{\delta} + 2\delta;$$

let us therefore assume otherwise, letting $p_{\hat{a}}$ be the larger value.

Let $W(\hat{a}) = \prod_{j=1}^{r} N_j(a_j)$ and similarly for $\hat{b}$; and let $W' = \prod_{j=3}^{r} N_j(a_j) = \prod_{j=3}^{r} N_j(b_j)$. From our assumption we have

$$\mathrm{E}(W(\hat{a}) - W(\hat{b})) \geqslant s^r(p_{\hat{a}} - \delta) - s^r(p_{\hat{b}} + \delta) \geqslant s^r(p_{\hat{a}} - p_{\hat{b}} - 2\delta) \geqslant 9s^r\sqrt{\delta}.$$

On the other hand,

$$\begin{aligned}
\mathrm{E}(W(\hat{a}) - W(\hat{b})) &= \mathrm{E}\left[(N_1(c)\,(N_2(d) - N_1(d)) + N_1(d)\,(N_1(c) - N_2(c)))\, W'\right] \\
&\leqslant \left[\mathrm{E}\left((N_1(c)W')^2\right)\right]^{1/2}\left[\mathrm{E}\left((N_2(d) - N_1(d))^2\right)\right]^{1/2} \\
&\quad + \left[\mathrm{E}\left((N_1(d)W')^2\right)\right]^{1/2}\left[\mathrm{E}\left((N_1(c) - N_2(c))^2\right)\right]^{1/2} \\
&\leqslant 2\left((s \cdot s^{r-2})^2\right)^{1/2} \cdot (8s)^{1/2} \\
&= \left(32s^{2r-1}\right)^{1/2}.
\end{aligned}$$

Comparing, we then have

$$32s^{2r-1} > 81s^{2r}\delta,$$

contradicting the choice of $s \geqslant 1/\delta$.

Since we can express any permutation $\pi \in S_r$ as a product of at most $r-1$ transpositions, we now have

$$\begin{aligned}
|p_{a_1,a_2,\ldots,a_r} - p_{a_{\pi(1)},\ldots,a_{\pi(r)}}| &< (r-1)\left(9\sqrt{\delta} + 2\delta\right) \\
&< 11r\sqrt{\delta} \\
&= \varepsilon/(2m^r)
\end{aligned}$$

so that *any* r-subsequence of the X_i^*s, in particular $X_1^*, \ldots, X_r^*$, lies within variation distance

$$m^r \cdot \varepsilon/(2m^r) = \varepsilon/2$$

of a precisely exchangeable sequence $Y_1, \ldots, Y_r$.

Since this means also that the variation distance between $X_1^*, \ldots, X_k^*$ and $Y_1, \ldots, Y_k$ is at most $\varepsilon/2$, we conclude that $X_1^*, \ldots, X_k^*$ is within variation distance at most $\varepsilon/2 + \varepsilon/2 = \varepsilon$ of an i.i.d. mix as required. □

Acknowledgments

The authors have benefited from valuable (and fascinating) conversations with David Aldous, Noga Alon and Persi Diaconis. We would also like to express our appreciation to an anonymous referee for a very careful reading and many helpful suggestions.

References

[1] Aldous, D. J. (1977) Subsequences of sequences of random variables. *Bull. Amer. Math. Soc.* **83** 121–123.

[2] Brightwell, G. R. and Scheinerman, E. R. (1992) Fractional dimension of partial orders. *Order* **9** 139–158.

[3] Chatterji, S. D. (1974) A principle of subsequences in probability. *Adv. Math.* **13** 31–54.

[4] Dacunha-Castelle, D. (1974) Indiscernability and exchangeability in L^p spaces. *Proc. Seminar on Random Series, Convex Sets and Geometry of Banach Spaces* Aarhus.

[5] Diaconis, P. and Freedman, D. (1980) Finite exchangeable sequences. *Ann. Prob.* **8** 745–764.

[6] Felsner, S. and Trotter, W. T. (1994) On the fractional dimension of partially ordered sets. *Discrete Math.* **136** 101–117.

[7] Figiel, T. and Sucheston, L. (1976) An application of Ramsey sets in analysis. *Adv. Math.* **20** 103–105.

[8] de Finetti, B. (1931) Funzione caratteristica de un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei*, Ser. 6, Memorie, Classe de Scienze Fisiche, Mathematiche e Naturali **4** 251–299.

[9] Fishburn, P. C. (1970) Intransitive indifference with unequal indifference intervals. *J. Math. Psych.* **7** 144–149.

[10] Füredi, Z., Hajnal, P., Rödl, V. and Trotter, W. T. (1991) Interval orders and shift graphs. In *Sets, Graphs and Numbers* (A. Hajnal and V. T. Sós, eds), Colloq. Math. Soc. Janos Bolyai **60** 297–313.

[11] Kingman, J. F. C. (1978) Uses of exchangeability (The 1977 Wald Memorial Lectures). *Ann. Probab.* **6** 183–197.

[12] Kleitman, D. J. and Markovsky, G. (1975) On Dedekind's problem: The number of isotone Boolean functions, II. *Trans. Amer. Math. Soc.* **213** 373–390.

[13] Komlós, J. (1967) A generalisation of a problem of Steinhaus. *Acta Math. Acad. Sci. Hungar.* **18** 217–229.

[14] Komlós, J. (1990) A strange pigeonhole principle. *Order* **7** 107–113.

[15] Trotter, W. T. (1992) *Combinatorics and Partially Ordered Sets: Dimension Theory*. The Johns Hopkins University Press, Baltimore, Maryland.

[16] Trotter, W. T. (1995) Partially ordered sets. In *Handbook of Combinatorics* (R. L. Graham, M. Grötschel and L. Lovász, eds), Vol. I, Elsevier, Amsterdam, pp. 433–480.

[17] Trotter, W. T. (1996) Graphs and partially ordered sets. *Congressus Numerantium* **116** 253–278.